J Prod Anal
DOI 10.1007/s11123-012-0316-9

Consistent estimation of technical and allocative efficiencies for a semiparametric stochastic cost frontier with shadow input prices

Tai-Hsin Huang · Kuan-Chen Chen ·
Chien-Hsiu Lin · Ming-Tai Chung

Abstract Conventional parametric stochastic cost frontier models are likely to suffer from biased inferences due to misspecification and the ignorance of allocative efficiency (AE). To fill up the gap in the literature, this article proposes a semiparametric stochastic cost frontier with shadow input prices that combines a parametric portion with a nonparametric portion and that allows for the presence of both technical efficiency (TE) and AE. The introduction of AE and the nonparametric function into the cost function complicates substantially the estimation procedure. We develop a new estimation procedure that leads to consistent estimators and valid TE and AE measures, which are proved by conducting Monte Carlo simulations.

Keywords Semiparametric cost frontier · Monte Carlo simulations · Shadow prices · Technical efficiency · Allocative efficiency

JEL Classification C14 · C15 · C33 · G21

T.-H. Huang (✉) · C.-H. Lin · M.-T. Chung
Department of Money and Banking, National Chengchi University, Taipei, Taiwan
e-mail: thuang@nccu.edu.tw

C.-H. Lin
e-mail: clin@nccu.edu.tw

M.-T. Chung
e-mail: 95352509@nccu.edu.tw

K.-C. Chen
School of Health Care Administration, Taipei Medical University, Taipei, Taiwan
e-mail: kcc@tmu.edu.tw

1 Introduction

A parametric linear or nonlinear regression model requires setting a specific functional form prior to estimation in order to describe the true but unknown relationship between the dependent and the independent variables. Consequently, potential specification errors are likely to occur, leading to an inconsistent estimation. Although some economic models do explicitly suggest relationships among economic variables, most implications of economic theory are nonparametric. Therefore, if one has reservations about a particular parametric form, then a nonparametric function can be an alternative candidate. Nonparametric regression models permit the functional relationship to be unknown and nevertheless fit the data quite well without imposing restrictions beyond some degree of smoothness. They deliver estimators and inference procedures that are less reliant on the imposition of specific functional forms. Inclusion of the nonparametric element may circumvent an inconsistent estimation arising from invalid parameterization. However, the inherent critical element of the "curse of dimensionality" limits the unknown function of a nonparametric model to contain a small number of variables to lessen the approximation error to the unknown function.

A researcher in some cases may be confident about a particular parametric form for one portion of the regression function, but less sure about the shape of another portion. Such prior beliefs justify the necessity for linking parametric with nonparametric techniques to formulate semiparametric regression models. The added value of semiparametric techniques consists in their competence to largely mitigate the

Published online: 22 September 2012

curse of dimensionality distress, and the respective estimators of the parametric and nonparametric components have their usual rates of convergence.[1] See, for example, Härdle (1990), Wand and Jones (1995), Fan et al. (1996), and Yatchew (1998, 2003).

Fan et al. (1996) first extended the traditional stochastic production frontier model, dated back to Aigner et al. (1977) and Meeusen and Van Den Broeck (1977), to a semiparametric frontier model in the context of cross section. They proposed pseudo-likelihood estimators and proved by Monte Carlo experiments that the finite-sample performance of their estimators is satisfactory. Deng and Huang (2008) further generalized it to a panel data setting and allowed for time-variant technical efficiency (TE) in the form of Battese and Coelli (1992). Nevertheless, almost all of the related works that use a semiparametric frontier model focus on the study of technical efficiency (TE) and the achievement of allocative efficiency (AE) is implicitly presumed. Kumbhakar and Wang (2006a) found that the assumption of fully AE in a cost function setting tends to bias parameter estimates of the cost function and subsequent measures using these estimates.

There are several ways of evaluating both TE and AE measures in the literature. The shadow input price approach will be used to estimate the shadow cost system, consisting of an expenditure (cost) equation and the corresponding share equations, simultaneously using the maximum likelihood. This allows researchers to decompose cost inefficiency into its technical and allocative components. Unfortunately, the highly nonlinear nature of the simultaneous equations makes the estimation almost untractable. Kumbhakar and Lovell (2000) proposed a two-step procedure with an eye to mitigate somewhat the estimation problem of a pure parametric shadow cost system. They recommend estimating the share equations in the first step by the method of nonlinear iterative seemingly unrelated regression (NISUR) to acquire the shadow price parameter estimates of interest.[2] These estimates are treated as given in the second step, where the maximum likelihood technique is exploited to estimate the stochastic cost frontier alone after appropriately transforming the original expenditure equation using the first step estimates. Kumbhakar and Lovell (2000) did not address the properties of their proposed estimators.[3]

The purpose of the current work is threefold. First, we relax the parametric restriction on a cost function representing technology in order to at least diminish the possible specification error. Second, the semiparametric stochastic shadow cost frontier offered by this paper differs from the standard semiparametric regression model and from the stochastic production frontier of Fan et al. (1996). Specifically, our model accommodates both TE and AE to avoid biased estimates of the technology parameters. To the best of our knowledge, no work has been done to introduce both efficiency measures into a semiparametric stochastic shadow cost frontier under the framework of panel data. It is hoped that this research will bridge the existing gap and to better characterize a firm's optimization behavior. Third, a distinct five-step procedure from the one suggested by Kumbhakar and Lovell (2000) is proposed to facilitate the estimation. We argue for the new procedure due to the fact that the resultant estimators are shown to converge to the true values as the sample size increases by applying Monte Carlo simulations.

The rest of this paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 first presents the semiparametric stochastic cost frontier with shadow input prices and then proposes the estimation procedure. Section 4 introduces the design of Monte Carlo experiments to be conducted in the next section. Section 5 provides and discusses the results of the experiments, which are intended to detect a suitable estimation procedure leading to consistent estimators, while the last section concludes the paper.

2 Literature review

The TE score of a firm can be estimated by two main approaches, i.e., data envelopment analysis (DEA) and stochastic frontier approach (SFA). The former involves mathematical programming without the need for specifying an explicit functional form, while the latter employs the econometric methods to deal with the composed random disturbances. These approaches have their own advantages and weaknesses. Fan et al. (1996) elegantly extended the standard parametric SFA to a semiparametric model in the context of cross section, where the functional form of the production frontier needs not to be specified a priori. Their method makes use of nonparametric regression techniques to avoid the requirement of specifying a particular production function, associating a firm's output with inputs.

[1] Robinson (1988) showed that the parametric estimators are consistent at the parametric rate of $N^{-1/2}$, while the nonparametric estimators converge at a slower rate than $N^{-1/2}$, where N denotes the sample size.

[2] It can be shown that these estimators are consistent and asymptotically normal.

[3] The two-step estimation procedures of Kumbhakar and Lovell (2000) are found to give consistent estimates of AE when the cost function takes the translog form. The TE estimates in general perform

Footnote 3 continued
not as well as AE estimates, due mainly to the badly performed distribution parameter estimates. Inferences on TE scores using a small sample are doubtful.

Therefore, the possible problem of misspecification is no longer a key issue as opposed to conventional parametric approach. Deng and Huang (2008) generalized the semiparametric model of Fan et al. (1996) to a panel data setting and allowed for time-varying TE. Their empirical evidence uncovered that the standard parametric translog production function tends to underestimate the TE score due to the possible specification error and its lack of flexibility in describing firms' production characteristics.

Wheelock and Wilson (2001) estimated and compared the measures of scale and scope economies for US commercial banks, derived from estimating parametric and nonparametric cost equations, without regard to TE and AE. In an expenditure equation modeling both technical inefficiency (TI) and allocative inefficiency (AI), it is difficult for researchers to appropriately relate the two-sided disturbances in the input share equations to the nonnegative AI term in the expenditure equation. This is known as the Greene problem (Bauer 1990). Berger et al. (1993), Atkinson and Cornwell (1994), Kumbhakar (1996a), Huang (2000), and Huang and Wang (2004), to mention a few, utilized shadow prices to account for AI in addition to TI. Kumbhakar (1996b, 1997) gave a complete treatment on how to model TI and AI concurrently. Kumbhakar and Wang (2006b) demonstrated an alternative primal system, consisting of a production function and the first-order conditions of cost minimization. However, the cost function associated with the translog production function cannot be analytically derived. The shadow price technique does not need to specify an ad hoc relationship between the AI term of the expenditure equation and the disturbance terms of the share equations. In addition, this technique can be applied to any parametric cost function as well as some semiparametric cost functions. We therefore adopt the technique throughout the paper.

3 Semiparametric stochastic shadow cost frontiers

Let the jth shadow input prices, W_j^*, be defined as

$$W_j^* = H_j W_j, \quad j = 1, \ldots, J \qquad (3-1)$$

where H_j (>0) denotes the allocative parameter of input j, measuring the extent to which the shadow and actual input prices (W_j) differ. The shadow price of W_j^* is not directly observable due to the presence of H_j that can be estimated. It thus reflects the degree of allocative inefficiency arising from, e.g., regulation or slow adjustment to changes in input prices. Here, a firm's decision is assumed to be grounded on shadow input prices. Following Atkinson and Cornwell (1994), Kumbhakar (1996b, 1997), Huang and Wang (2004), and Huang et al. (2011), the minimized efficiency adjusted shadow cost, C^{**}, for a firm employing input vector X to produce output vector Y can be expressed as[4]:

$$\begin{aligned} C^{**}\left(Y, \frac{W^*}{b}\right) &= \underset{bX}{\mathrm{Min}}\left[\frac{W^*}{b}(bX) \middle| F(bX, Y) = 0\right] \\ &= \frac{1}{b} C^*(Y, W^*) \end{aligned} \qquad (3-2)$$

where $b(0 < b \leq 1)$ represents the degree of input-oriented TI, C^* is referred to as the shadow cost function independent of the TI parameter of b, and Y is an m-vector of output quantities. A firm is said to be technically efficient if it has a value of $b = 1$, while a firm operating beneath the efficiency frontier has a value of $b < 1$. The larger the value of b is, the more technically efficient the firm will be. Function $F(\cdot,\cdot)$ represents the production transformation function with input-oriented technical inefficiency (Atkinson and Cornwell 1993). Note that b is an unknown parameter to be estimated later.

Since a cost function must satisfy the homogeneity restriction of degree one in input prices, we can only measure $J - 1$ relative allocative parameters H_j/H_J, $j = 1, \ldots, J - 1$, and the Jth input is arbitrarily chosen as the numeraire. A value of H_j/H_J less (greater) than unity means that the jth input tends to be overused (underused) relative to input J. Either overuse or underuse reflects the presence of AI. Using Shephard's Lemma, the shadow cost share equation of input j ($j = 1,\ldots, J - 1$) is written as:

$$S_j^*(W^*, Y) \equiv \frac{\partial \ln C^*}{\partial \ln W_j^*} = \frac{bW_j^* X_j}{C^*}. \qquad (3-3)$$

After some manipulations and taking a natural logarithm, a firm's actual expenditure (E) can be linked with C^{**} (C^*) and S_j^* as follows:

$$\begin{aligned} \ln E &= \ln C^{**} + \ln \sum_j H_j^{-1} S_j^* \\ &= \ln C^*(Y, W^*) + \ln \sum_j H_j^{-1} S_j^* + U \end{aligned} \qquad (3-4)$$

where $U = -\ln b$ reflects the additional (log) expenditure incurred by TI and is specified as a one-sided error term later, $\ln \sum_j H_j^{-1} S_j^*$ captures a partial extra cost entailed by AI, and the remaining extra cost of AI is embedded in ln $C^*(Y, W^*)$ due to $W^* \neq W$. It is readily seen that the extra cost arising from AI vanishes once the firm attains fully AE.

Equation (3-4) becomes a regression equation after appending a two-sided random disturbance v to it, where v is assumed to be distributed as $N(0, \sigma_v^2)$. Term $U + v$ forms

[4] Note that the objective function of (3-2) is initially expressed as W^*X and the choice vector is X. Since parameter b emerges in the constraint of $F(\cdot,\cdot)$, together with X, it is convenient to transform the objective function into $(W^*/b)bX$. This is equivalent to treat W^*/b as the new input prices and bX the new choice vector.

the composed error term. This equation associates TI with AI systematically. To identify the allocative parameters, one has to count on the share equations. It can be shown that the actual share equation of input j (S_j) is formulated as

$$S_j = \frac{H_j^{-1} S_j^*}{\sum_j H_j^{-1} S_j^*}, \quad j = 1, \ldots, J. \qquad (3-5)$$

After adding random disturbances to these share equations, they can be used to help estimate parameters H_j. When panel data are available, it is more ambitious to assume TI term U to be time-varying. The time-variant TE model of Battese and Coelli (1992) is adopted with $U_{nt} = u_n \exp[-\gamma(t-T)]$, $n = 1, \ldots, N$, $t = 1, \ldots, T$, where u_n is a firm-specific TI random variable distributed as $|N(0, \sigma_u^2)|$ and independent of v_{nt}, and $g(t) = \exp[-\gamma(t-T)]$ contains an extra parameter γ to be estimated.[5]

We now turn to the functional form of $\ln C^*(Y, W^*)$ in (3-4). It is conventionally specified as a translog form, or as a Fuss functional form like Berger et al. (1993), or as a Fourier flexible function such as Altunbas et al. (2001) and Huang and Wang (2004). In this paper $\ln C^*(Y, W^*)$ is instead formulated as a semiparametric form:

$$\ln C^*(Y_{nt}, W_{nt}^*) = X_{nt}\beta + M(\ln Y_{nt}) \qquad (3-6)$$

where X_{nt} consists of the linear and quadratic terms of $\ln W_{jnt}^*$ ($j = 1, \ldots, J$), the cross product terms among $\ln W_{jnt}^*$, and the cross product terms of $\ln W_{jnt}^*$ with $\ln Y_{int}$ ($i = 1, \ldots, m$), β is the corresponding unknown parameter vector, $\ln Y_{nt}$ is a $m \times 1$ random vector of (log) outputs with support, and $M(\cdot)$ is assumed to be a smooth function with unknown form. The reason why the non-parametric part has only outputs is that a cost function should be homogeneous of degree one in input prices. If the non-parametric part contains input prices, then it is difficult to impose this restriction on it. Our specification is similar to, e.g., Akhigbe and McNulty (2003), Altunbas et al. (2001), Berger and DeYoung (1997), Berger et al. (1997), and Berger and Mester (1997), who applied the Fourier flexible (FF) cost function with the terms of Fourier series being composed of transformed outputs only, excluding input prices, for the same reasoning.

We rewrite our cost function system as:

$$\ln E_{nt} = X_{nt}\beta + \ln G_{nt} + M(\ln Y_{nt}) + \varepsilon_{nt} \qquad (3-7)$$

$$S_{jnt} = \frac{H_j^{-1} S_{jnt}^*}{\sum_j H_j^{-1} S_{jnt}^*} + \eta_{jnt}, \quad j = 1, \ldots, J \qquad (3-8)$$

where $\ln G_{nt} = \ln \sum_j H_j^{-1} S_j^*$, $\varepsilon_{nt} = u_n \exp[-\gamma(t-T)] + v_{nt}$, and $\eta_{nt} = (\eta_{1nt}, \ldots, \eta_{Jnt})'$ is assumed to be a multivariate normal random vector with mean zero and constant covariance matrix, independent of ε_{nt}. It can be shown that the nth firm's probability density function of the composed disturbance $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nT})'$ is equal to:

$$h(\varepsilon_n) = \frac{2}{\sigma_v^{T-1}\sigma}[1 - \Phi(A_n)]\left[\prod_{t=1}^{T} \phi\left(\frac{\varepsilon_{nt}}{\sigma_v}\right)\right] \exp\left(\frac{1}{2}(-A_n)^2\right) \qquad (3-9)$$

where $A_n = -\lambda \sum_t \varepsilon_{nt} g_t / \sigma, g_t = e^{-\gamma(t-T)}, \lambda = \sigma_u/\sigma_v, \sigma^2 = \sigma_v^2 + \sigma_u^2 \sum_{t=1}^{T} g_t^2$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and standard normal cumulative distribution functions, respectively. The log-likelihood function of expenditure Eq. (3-7) alone can be easily derived by first multiplying (3-9) over firms and then taking the natural logarithm. Combining (3-9) with the joint probability density function of the $(J-1)$ random disturbances of the share equations (η_{nt}), the cost function system can be simultaneously estimated by the maximum likelihood if M has a known form. Readers are suggested to refer to, e.g., Ferrier and Lovell (1990) and Kumbhakar (1991), for details.

Three difficulties deserve specific mention. First, since the log-likelihood function of the above cost function system is highly nonlinear, getting maximum likelihood estimators is computationally difficult, even though not infeasible. Second, M has an unknown functional form, hindering the log-likelihood function of the expenditure equation from being maximized with respect to M in particular. One alternative relies on the use of some non-parametric approaches to estimate M. However, M cannot be estimated directly by existing nonparametric regression methods, because it is not the conditional expectation of $\ln E_{nt} - X_{nt}\beta - \ln G_{nt}$ given $\ln Y_{nt}$. This is caused by the nonzero mean of one-sided error U_{nt}, i.e.:

$$\begin{aligned} E(\ln E - X\beta - \ln G | \ln Y_{nt}) &= M(\ln Y_{nt}) + \mu_t(\sigma^2, \lambda, \gamma) \\ &\neq M(\ln Y_{nt}) \end{aligned} \qquad (3-10)$$

where

$$\begin{aligned} \mu_t(\sigma^2, \lambda, \gamma) = E(U_{nt}) &= g_t \frac{\sqrt{2}}{\sqrt{\pi}} \sigma_u \\ &= g_t \frac{\sqrt{2}}{\sqrt{\pi}} \left(\frac{\lambda\sigma}{\sqrt{1 + \lambda^2 \sum_t g_t^2}} \right). \end{aligned} \qquad (3-11)$$

[5] Term g_t decreases at an increasing rate if $\gamma > 0$, increases at an increasing rate if $\gamma < 0$, or stays constant if $\gamma = 0$.

One cannot separate $M(\ln\ Y_{nt})$ from $E(\ln\ E - X\beta - \ln G|\ln Y_{nt})$ of (3-10) by employing a nonparametric estimation. This problem can be solved by substituting $E(\ln E - X\beta - \ln G|\ln Y_{nt}) - \mu_t$ for $M(\ln Y_{nt})$into the log-likelihood function. $E(\ln E - X\beta - \ln G|\ln Y_{nt})$ can now be consistently estimated by the nonparametric approach. For details, please see, e.g., Fan et al. (1996) and Deng and Huang (2008). Finally, term $\ln G_{nt} = \ln \sum_j H_j^{-1} S_j^*$ is obviously a nonlinear function of unknown parameters, leading the kernel estimation procedure for a standard semiparametric regression model, as proposed by Robinson (1988), to be not applicable. We shall discuss possible ways of getting rid of this difficulty in Subsect. 4-1, which influence the consistency of the parameter estimates and are the core of this study.

We adopt the kernel estimation technique to estimate the conditional expectations, such as $E(\ln E|\ln Y_{nt})$, since it is one of the popular nonparametric estimation methods. Specifically, the Nadaraya-Watson kernel estimator (Nadaraya 1964; Watson 1964) for a scalar $\ln Y_{nt}$ is given by:

$$\hat{E}(\ln E|\ln Y_{nt}) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \ln E_{it} K\left(\frac{\ln Y_{nt} - \ln Y_{it}}{h}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T} K\left(\frac{\ln Y_{nt} - \ln Y_{it}}{h}\right)} \qquad (3-12)$$

where $K(\cdot)$ is the kernel function and h is the smoothing parameter. Equation (3-12) can be easily extended to a higher dimensional case of $\ln Y_{nt}$. The rest of the conditional expectations in Eq. (3-10) can be estimated analogously.

We now outline the estimation procedure of the semi-parametric shadow cost frontier in the following five steps.

Step 1 Simultaneously estimate the $J - 1$ input share equations of (3-8) by the NISUR to obtain the $J - 1$ estimates of relative allocative parameters H_j/H_J $(j = 1, \ldots, J - 1)$ and a part of the parameters involving the input prices of expenditure Eq. (3-7). These estimates can be shown to be consistent and are used to calculate $\ln G_{nt}$, denoted by $\ln \hat{G}_{nt}$.

Step 2 Apply formula (3-12) to obtain the kernel estimates of $E(\ln E|\ln Y_{nt}.)$, $E(X|\ln Y_{nt}.)$, and $E(\ln \hat{G}|\ln Y_{nt})$, denoted by $\hat{E}(\ln E|\ln Y_{nt})$, $\hat{E}(X|\ln Y_{nt})$, and $\hat{E}(\ln \hat{G}|\ln Y_{nt})$, respectively.

Step 3 Equation (3-7) subtracts its own conditional expectations on $\ln Y_{nt}$ to yield

$$\ln E_{nt} - E(\ln E|\ln Y_{nt}) = [X_{nt} - E(X|\ln Y_{nt})]\beta + \ln G_{nt} - E(\ln \hat{G}|\ln Y_{nt}) + \varepsilon'_{nt}. \qquad (3-13)$$

After substituting the kernel estimates derived in Step 2 for those conditional expectations in (3-13),[6] parameters β can be consistently estimated by the nonlinear least squares method, since the new error component $\varepsilon'_{nt} (= v_{nt} + U_{nt} - \mu_t)$ has zero mean asymptotically. The nonlinear least squares is required due to the nonlinearity of $\ln G_{nt}$. This distinguishes the current paper from Robinson (1988), where the ordinary least squares apply.

Step 4 Let

$$\hat{\varepsilon}_{nt} = \ln E_{nt} - \hat{E}(\ln E|\ln Y_{nt}) - [X_{nt} - \hat{E}(X|\ln Y_{nt})]\hat{\beta} - \ln \hat{G}_{nt} + \hat{E}(\ln \hat{G}|\ln Y_{nt}) + \mu_t \qquad (3-14)$$

Maximizing the log-likelihood function derived from (3-9) with ε_{nt} replaced by $\hat{\varepsilon}_{nt}$ over σ^2 and λ, one obtains the solution to σ after tedious manipulation as

$$\hat{\sigma} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (3-15)$$

where $a = 1 - 2\lambda^2 \sum_t g_t^2/(\pi TT)$, $TT = T + \lambda^2 \sum_t g_t^2$,

$$b = -2^{3/2}\lambda\sqrt{(1 + \lambda^2 \sum_t g_t^2)/\pi} \sum_i \sum_t e_{it} g_t/(nTT),$$

$$e_{it} = \ln E_{nt} - \hat{E}(\ln E|\ln Y_{nt}) - [X_{nt} - \hat{E}(X|\ln Y_{nt})]\ \hat{\beta} - [\ln \hat{G}_{nt} - \hat{E}\ (\ln G|\ln Y_{nt})],$$

and

$c = -(1 + \lambda^2 \sum_t g_t^2) \sum_i \sum_t e_{it}^2/(nTT)$.

In (3-15) notation "^" is added on σ since the kernel and NISUR estimators of $\hat{E}(.|x_{it})$ and $\hat{\beta}$ are used to replace their respective true counterparts. For details, please see Deng and Huang (2008) for a panel data setting with time variant TI. Because $\hat{\sigma}$ is a function of λ, γ, and data, it can be concentrated out of the log-likelihood function to reduce the number of unknown parameters.

Step 5 Maximize the concentrated log-likelihood function of the expenditure equation over the remaining two unknown parameters of λ and γ, where ε_{nt} is replaced by $\hat{\varepsilon}_{nt}$ in Step 4. The resulting pseudolikelihood estimates are denoted by $\hat{\lambda}$ and $\hat{\gamma}$. Substituting them into (3-15), we get the estimate of σ and still signify it by $\hat{\sigma}$. Plugging the three estimates into (3-11) yields the estimate of μ_t, denoted by $\hat{\mu}_t$. Finally, the nonparametric function $M(\ln Y_{nt})$ can be consistently estimated by

$$\hat{M}(\ln Y_{nt}) = \hat{E}(\ln E|\ln Y_{nt}) - \hat{E}(X|\ln Y_{nt})\hat{\beta} - \hat{E}(\ln G|\ln Y_{nt}) - \hat{\mu}_t \qquad (3-16)$$

[6] There is a concern with the referee's suggestion that equation (3-13) would be plus an extra term $E(\ln \hat{G}_{nt}|\ln Y_{nt}) - E(\ln G_{nt}|\ln Y_{nt})$, which is non-zero. We can examine how fast it converges to zero as either N or T grows by simulations. The results reveal that the bias measures are small in all of the (N, T) combinations. In addition, the biases decrease as either N or T grows, with the exception of the case of $(N, T) = (50, 20)$. This leads us to conclude that the extra term $E(\ln \hat{G}_{nt}|\ln Y_{nt}) - E(\ln G_{nt}|\ln Y_{nt})$ does converge rapidly.

where $\hat{\beta}$ comes from the estimates of Step 3.It is well known that the maximum likelihood estimator of λ and γ must be asymptotically unbiased and efficient if the regularity conditions hold. Although the individual kernel regression estimators of Step 2 have pointwise convergence rates slower than root-NT ($NT^{-1/2}$), where NT signifies the sample size, the average quantities of the elements in (3-15) have an order of $O_p(NT^{-1/2})$ under very weak conditions. See, for example, Härdle and Stoker (1989) and Fan and Li (1992). Fan et al. (1996) claimed that $\hat{\sigma}^2 - \sigma^2 = O_p(NT^{-1/2})$ under quite weak conditions. As estimator $\hat{M}(\ln Y_{nt})$ of (3-16) is a function of several kernel regression estimators, having slower convergence rates than $NT^{-1/2}$, it consequently converges to $M(\ln Y_{nt})$ for each nt at a slower rate than $NT^{-1/2}$.

The foregoing five steps constitute the entire estimation procedure and the resulting estimates can be further utilized to evaluate, e.g., measures of AE and TE. In particular, the formula proposed by Battese and Coelli (1992) is adopted to gauge each firm's TE score. Based on (3-4), the (log) cost of AI, denoted by u_{nt}^{AI}, is defined as the difference between the (log) shadow expenditure ($\ln C^*(Y, W^*) + \ln \sum_j H_j^{-1} S_j^*$) and the (log) optimized cost ($\ln C(Y, W)$) that achieves AE, i.e.:

$$u_{nt}^{AI} = \ln C^*\left(Y_{nt}, W_{nt}^*\right) + \ln G\left(Y_{nt}, W_{nt}^*\right) - \ln C(Y_{nt}, W_{nt}), \tag{3 − 17}$$

which is a non-negative value by definition. The measure of AE is then obtained by taking the natural exponent of $-u_{nt}^{AI}$, which ranges from zero to unity.

There are three attributes worth noting. First, the consistent estimates of $J - 1$ relative allocative parameters H_j/H_J ($j = 1,\ldots, J - 1$) yielded in Step 1 are treated as given throughout the remaining four steps. This avoids estimating the whole cost system simultaneously by the maximum likelihood and the difficulty in achieving convergence, on the one hand. The number of parameters to be estimated in later steps is largely decreased, on the other hand. Second, despite the fact that $\ln \hat{G}_{nt}$ can be computed in Step 1 and is used to obtain kernel estimate $\hat{E}(\ln \hat{G}|\ln Y_{nt})$ in Step 2, parameters included in $\ln G_{nt}$ of (3-13) need to be estimated again along with β, even though they have already been estimated in Step 1. Conversely, Kumbhakar and Lovell (2000) suggested subtracting $\ln \hat{G}_{nt}$ directly from the dependent variable of (3-13), which may give rise to undesirable estimation results. We will come back to this in Sect. 5. Third, since Step 5 aims to estimate merely λ and γ, the log-likelihood function is usually not very difficult to converge.

4 Monte Carlo simulations

This section first proposes three models to be used to compare the performance of their estimators. The next subsection specifies an expenditure equation and addresses the data generation processes for all variables involved.

4.1 Design of experiments

We plan to perform Monte Carlo simulations using three models and evaluate the performance of their estimates in terms of bias and mean square errors (MSE). Model A follows the five steps addressed by the previous section. Models B and C are simplified version of Model A for the purpose of making comparisons among the three models. At the outset, all of the three models have to estimate the input share equations using the NISUR, i.e., carrying out the first step. The $J - 1$ allocative parameter estimates are next applied to estimate $\ln G_{nt}$ and AE, denoted by G1 and AE1, respectively, while the subsequent steps of the three models differ from one another. Note that the $J - 1$ allocative parameters are treated as given thereafter. The three models are as follows.

4.1.1 Model A

This preferred model follows exactly the above five steps. Using the kernel estimates of $\hat{E}(\ln E|\ln Y_{nt})$, $\hat{E}(X|\ln Y_{nt})$, and $\hat{E}(\ln \hat{G}|\ln Y_{nt})$ from Step 2, we estimate Eq. (3-13) by the NISUR in Step 3 to obtain the estimates of β. At the same time, nonlinear function $\ln G_{nt}$ is assumed to be unknown, i.e., all of the parameters shown in the parametric part of the cost function are jointly estimated, but exclude the parameters associated with the distribution of v and U. Estimates $\hat{\beta}$ together with the $J - 1$ allocative parameter estimates are employed to calculate new estimates of $\ln G_{nt}$ and AE, denoted by G2A and AE2A. The remaining parameters embedded in the distributions of v and U are estimated in Steps 4 and 5.

4.1.2 Model B

This model is similar to Model A except that function $\ln G_{nt}$ is treated in a different way from Model A. Specifically, the estimated $\ln G_{nt}$, $\ln \hat{G}_{nt}$, derived from Step 1 is viewed as fixed so that it can be subtracted from the dependent variable. The new transformed equation becomes

$$\begin{aligned} &\ln E_{nt} - \ln \hat{G}_{nt} - E(\ln E - \ln \hat{G}|\ln Y_{nt}) \\ &\quad = [X_{nt} - E(X|\ln Y_{nt})]\beta + \hat{\varepsilon}'_{nt} \end{aligned} \tag{4 − 1}$$

where the notations are similarly defined to (3-13). After substituting the kernel estimates of $\hat{E}(\ln E - \ln \hat{G}|\ln Y_{nt})$ and $\hat{E}(X|\ln Y_{nt})$ for the corresponding conditional means in (4-1), β is estimated simply by ordinary least squares (OLS). This procedure is analogous to the one proposed by Kumbhakar and Lovell (2000, pp. 295–296) in spirit, while their underlying model is parametric. Estimates $\hat{\beta}$ are next used to compute $\ln G_{nt}$ and AE, denoted by G2B and AE2B. Finally, Steps 4 and 5 are executed.

4.1.3 Model C

This model is further simplified from Model B and is similar to the one suggested by Kumbhakar and Lovell (2000, p. 165) in essence. Again, their underlying model is parametric. Since the input share equations include vector β, their consistent estimate $\hat{\beta}$ from Step 1 can be treated as fixed. In this manner, the new dependent variable turns out to be $\ln E_{nt} - \ln \hat{G}_{nt} - X_{nt}\hat{\beta}$ with corresponding kernel estimate $\hat{E}(\ln E - \ln \hat{G} - X\hat{\beta}|\ln Y_{nt})$ obtained by Step 2. Step 3 is no longer needed and Eq. (3-14) of Step 4 is modified accordingly as:

$$\begin{aligned}\hat{\varepsilon}_{nt} &= \ln E_{nt} - \ln \hat{G}_{nt} - X_{nt}\hat{\beta} \\ &\quad - \hat{E}(\ln E - \ln \hat{G} - X\hat{\beta}|\ln Y_{nt}) + \mu_t\end{aligned} \quad (4-2)$$

After concentrating out σ^2, we execute Step 5. This completes the entire procedure.

It is seen that the major differences among the three models stem from distinct treatments on $\ln \hat{G}_{nt}$ and $\hat{\beta}$. As a result, we can compare the performance of the resulting estimates among these models, including the distribution parameters of v and U.

4.2 Model specifications

This subsection specifies the expenditure equation and the data generation processes for all variables involved that will be used to carry out Monte Carlo simulations to investigate the finite-sample performance of the proposed estimators in the last subsection. Since we are also interested in the effects of the number of firms (N) and time periods (T) on the parameter estimates, we consider several (N, T) combinations. Specifically, we choose $N = 50$, 100, 200 with $T = 6$, 10, 20.[7] Following Olson et al. (1980) and Fan et al. (1996), we consider three sets of variances and variance ratios, viz. $(\sigma^2, \lambda) = (1.88, 1.66)$, $(1.35, 0.83)$, $(1.63, 1.24)$. Finally, $\gamma = \pm\, 0.025$ are arbitrarily chosen.

The semiparametric cost frontier incorporating a single output and three inputs is formulated as:

$$\begin{aligned}\ln E &= M(\ln Y) + X\beta + \ln G + u + v \\ &= 2\ln(1 + y_1) + b_2 \ln(W_2^*) + b_3 \ln(W_3^*) \\ &\quad + \frac{1}{2} d_{22}[\ln(W_2^*)]^2 + \frac{1}{2} d_{33}[\ln(W_3^*)]^2 \\ &\quad + d_{23} \ln(W_2^*) \ln(W_3^*) + e_{12} \ln y_1 \ln(W_2^*) \\ &\quad + e_{13} \ln y_1 \ln(W_3^*) + \ln G + u + v\end{aligned} \quad (4-3)$$

Here, smooth function $M(\cdot)$ is arbitrarily assumed to be equal to $2\ln(1 + y_1)$. Recall that a cost function is required to be linearly homogeneous in input prices and symmetrical by the microeconomic theory. We randomly pick W_1 to normalize dependent variable E and the other two input prices to satisfy this requirement. The symmetry restriction is already imposed on (4-3).[8] To understand whether the performance of the estimates is robust to changes in the functional form of $M(\cdot)$, we specify an alternative form of $M(\cdot) = 0.2y_1$. We also extend (4-3) to a two-output and three-input case, assuming either $M(\cdot) = 2\ln(1 + y_1) + \ln y_2$ or $M(\cdot) = 2\ln y_2 + \sqrt{y_1 y_2}$. Note that in this extended case, the parametric part of (4-3) has to contain extra terms involving the cross products of $\ln y_2$ and (log) normalized input prices.

Input prices W_1, W_2, and W_3 are randomly drawn from dissimilar uniform distributions $U(0, 1)$, $U(0, 0.5)$, and $U(0.5, 1)$, respectively. The three-input and two-output quantities of x_1, x_2, x_3, y_1, and y_2 are independently generated from normal distributions $N(5, 0.5)$, $N(3, 0.1)$, $N(5, 0.5)$, $N(31, 10.1)$, and $N(20, 0.8)$, respectively. Two-sided error v is drawn from $N(0, \sigma_v^2)$ and one-sided error u from a half-normal $N^+(0, \sigma_u^2)$. The simulations are executed 1000 times for each model and the bias and the MSE are computed based on the 1000 replications. We set $H_2/H_1 = 0.8$ and $H_3/H_1 = 1.2$. The true values of the coefficients are as follows: $b_2 = 0.3$, $b_3 = 0.7$, $d_{22} = -0.05$, $d_{33} = -0.02$, $d_{23} = 0.5$, $e_{12} = 0.7$, $e_{13} = 0.9$, $e_{22} = 0.3$, and $e_{23} = 0.5$.

The corresponding input share equations can be deduced by taking the first partial derivatives of $\ln E$ with respect to $\ln W_i$, $i = 1, 2, 3$. Although the functional form of an expenditure equation is not unique, we recommend using those such as (4-3). A prominent feature of (4-3) consists in its smooth function being specified as a function of (log) outputs only, i.e., the (shadow) input prices are excluded. Otherwise, one is confronted with a problem on how to

[7] To save space, the results for the case of N = 50 are not shown, but available upon request from the authors.

[8] We also check whether the other two regularity conditions are satisfied, that is, a cost function is concave in input prices and the marginal cost should be positive. The model now is specified with an output (y) and two inputs (w_1, w_2) for simplicity. The result presents that most of the simulated outcomes meet the requirements, although the last condition performs a little worse for smaller sample. We conclude that vast majority of the simulated results satisfy the regularity properties.

properly disentangle the allocative parameters contained in $M(\cdot)$. More importantly, the share equations are unable to be explicitly derived by taking partial derivatives due to the unknown smooth function dependent of shadow prices. This impedes a researcher from subsequently identifying the allocative parameters.

5 Simulation results

This section compares the performance of the estimators discussed in the previous section. Table 1 summarizes the simulation outcomes of the empirical moments, i.e., bias and MSE, from the estimators for the nine (N, T) combinations. Let's first look at the performance of the allocative parameters, estimated in the first step. One thing that is immediately noticeable is that H_2/H_1 and H_3/H_1 are well estimated even for the case of the smallest sample size, i.e., (N, T) = (50, 6). Another desirable feature is that the bias and the MSE fall when either N or T increases, aside from the bias of H_2/H_1 when $N = 200$. Even in those exceptional cases the biases are negligible.

Table 1 The performance of the allocative parameter estimates setting $M(\cdot) = 2\ \ln(1 + y_1)$

	H_2/H_1		H_3/H_1	
(N, T)	Bias	MSE	Bias	MSE
(50, 6)	0.0006	0.0013	0.0048	0.0019
(50, 10)	0.0004	0.0007	0.0030	0.0010
(50, 20)	0.0001	0.0004	0.0026	0.0005
(100, 6)	0.0002	0.0006	0.0029	0.0009
(100, 10)	0.0001	0.0004	0.0026	0.0005
(100, 20)	−0.0001	0.0002	0.0010	0.0002
(200, 6)	−1.74E-06	0.0003	0.0021	0.0004
(200, 10)	−0.0001	0.0002	0.0010	0.0002
(200, 20)	0.0003	0.0001	0.0004	0.0001

Table 2 reveals that in general the MSEs of the parameter estimates of the parametric portion fall quickly as either N or T grows. For instance, when fixing $N = 50$, the MSE of the coefficient of $\ln(w_3/w_1)$ shrinks swiftly from 0.2227 to 0.0579 (not shown) as T increases from 6 to 20. The figure continues to fall to 0.0142 when (N, T) = (200, 20). In addition, the bias measures exhibit a similar pattern. In summary, the estimators in the first step work quite well as expected in terms of their biases and MSEs, which improve with the increase in either N or T.

Table 3 presents the biases and MSEs of the parametric part for Models A and B obtained from Step 3. Generally speaking, these estimators perform poorly. Their biases and MSEs are much larger than those derived from the first-stage estimation. In addition, the biases and MSEs of Model A decrease to some extent as the sample size increases, while the biases of Model B are hardly altered with the increase in the sample size. This leads us to conclude that the computationally simple first-stage estimators of the parametric part outperform the third-step

Table 2 The performance of the parameter estimates in Step 1 setting $M(\cdot) = 2\ \ln(1 + y_1)$

(N, T)	(100, 6)		(100, 10)		(100, 20)	
	Bias	MSE	Bias	MSE	Bias	MSE
$\ln(w_3/w_1)$	−0.0179	0.0935	−0.0055	0.0579	−0.0082	0.0284
$[\ln(w_3/w_1)]^2$	0.0036	0.0025	0.0025	0.0014	0.0014	0.0007
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	0.0015	0.0052	−0.0007	0.0036	0.0008	0.0016
$\ln y_1\ \ln(w_3/w_1)$	0.0046	0.0276	−0.0002	0.0178	0.0026	0.0083
$\ln(w_2/w_1)$	−0.0171	0.0679	−0.0049	0.0427	−0.0075	0.0211
$[\ln(w_2/w_1)]^2$	0.0011	0.0012	0.0010	0.0007	0.0006	0.0003
$\ln y_1\ \ln(w_2/w_1)$	0.0023	0.0157	−0.0016	0.0105	0.0015	0.0050
(N, T)	(200, 6)		(200, 10)		(200, 20)	
	Bias	MSE	Bias	MSE	Bias	MSE
$\ln(w_3/w_1)$	−0.0084	0.0489	−0.0082	0.0284	−0.0008	0.0142
$[\ln(w_3/w_1)]^2$	0.0026	0.0012	0.0014	0.0007	−0.0005	0.0003
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	−0.0004	0.0029	0.0008	0.0016	0.0003	0.0008
$\ln y_1\ \ln(w_3/w_1)$	0.0014	0.0148	0.0026	0.0083	−0.0009	0.0042
$\ln(w_2/w_1)$	−0.0033	0.0354	−0.0075	0.0211	−0.0022	0.0105
$[\ln(w_2/w_1)]^2$	0.0011	0.0006	0.0006	0.0003	−0.0001	0.0002
$\ln y_1\ \ln(w_2/w_1)$	−0.0013	0.0087	0.0015	0.0050	−0.0003	0.0025

Table 3 The performance of the parameter estimates from the third-stage setting $M(\cdot) = 2\ \ln(1 + y_1)$

(σ^2, λ)	(1.88, 1.66)				(1.35, 0.83)				(1.63, 1.24)			
	Model A		Model B		Model A		Model B		Model A		Model B	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(N, T) = (100, 6)												
$\ln(w_3/w_1)$	0.8227	0.7571	−0.3908	0.4784	0.8286	0.8027	−0.3999	0.6393	0.8250	0.7712	−0.3944	0.5251
$[\ln(w_3/w_1)]^2$	−0.0078	0.0404	0.7216	0.8238	−0.0089	0.0500	0.7323	0.9832	−0.0081	0.0433	0.7258	0.8709
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	0.0083	0.0400	−0.5468	0.3606	0.0094	0.0482	−0.5441	0.3881	0.0085	0.0425	−0.5458	0.3682
$\ln y_1\ \ln(w_3/w_1)$	−0.2446	0.0661	−0.9165	0.8992	−0.2462	0.0697	−0.9139	0.9232	−0.2452	0.0672	−0.9155	0.9056
$\ln(w_2/w_1)$	−0.8548	0.8141	0.1972	0.0970	−0.8587	0.8568	0.1945	0.1242	−0.8563	0.8271	0.1961	0.1048
$[\ln(w_2/w_1)]^2$	−0.0079	0.0440	0.7469	0.5831	−0.0091	0.0531	0.7501	0.5998	−0.0082	0.0467	0.7482	0.5883
$\ln y_1\ \ln(w_2/w_1)$	0.2536	0.0709	0.1990	0.0630	0.2548	0.0743	0.1953	0.0724	0.2541	0.0719	0.1975	0.0655
(N, T) = (100, 10)												
$\ln(w_3/w_1)$	0.6792	0.4909	−0.3976	0.2700	0.6828	0.5104	−0.3997	0.3295	0.6806	0.4967	−0.3984	0.2861
$[\ln(w_3/w_1)]^2$	−0.0019	0.0186	0.7234	0.6283	−0.0022	0.0221	0.7275	0.6892	−0.0019	0.0196	0.7250	0.6452
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	0.0028	0.0188	−0.5493	0.3237	0.0029	0.0220	−0.5464	0.3312	0.0028	0.0197	−0.5482	0.3253
$\ln y_1\ \ln(w_3/w_1)$	−0.2019	0.0431	−0.9174	0.8632	−0.2029	0.0447	−0.9147	0.8688	−0.2023	0.0436	−0.9164	0.8642
$\ln(w_2/w_1)$	−0.7082	0.5319	0.1985	0.0602	−0.7098	0.5489	0.1956	0.0692	−0.7089	0.5368	0.1974	0.0625
$[\ln(w_2/w_1)]^2$	−0.0035	0.0208	0.7490	0.5698	−0.0036	0.0245	0.7501	0.5763	−0.0034	0.0218	0.7494	0.5718
$\ln y_1\ \ln(w_2/w_1)$	0.2103	0.0466	0.1987	0.0477	0.2108	0.0480	0.1970	0.0515	0.2105	0.0470	0.1980	0.0486
(N, T) = (100, 20)												
$\ln(w_3/w_1)$	0.5231	0.2817	−0.3974	0.1864	0.5242	0.2861	−0.3972	0.1949	0.5235	0.2829	−0.3974	0.1864
$[\ln(w_3/w_1)]^2$	0.0009	0.0076	0.7203	0.5472	0.0013	0.0081	0.7209	0.5566	0.0011	0.0077	0.7203	0.5472
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	−0.0017	0.0076	−0.5471	0.3047	−0.0020	0.0081	−0.5461	0.3052	−0.0018	0.0077	−0.5471	0.3047
$\ln y_1\ \ln(w_3/w_1)$	−0.1552	0.0247	−0.9181	0.8483	−0.1556	0.0251	−0.9170	0.8477	−0.1554	0.0248	−0.9181	0.8483
$\ln(w_2/w_1)$	−0.5457	0.3055	0.1975	0.0441	−0.5458	0.3085	0.1964	0.0452	−0.5458	0.3063	0.1975	0.0441
$[\ln(w_2/w_1)]^2$	0.0021	0.0081	0.7496	0.5642	0.0023	0.0087	0.7498	0.5651	0.0022	0.0083	0.7496	0.5642
$\ln y_1\ \ln(w_2/w_1)$	0.1622	0.0269	0.1997	0.0421	0.1623	0.0272	0.1993	0.0426	0.1623	0.0270	0.1997	0.0421
(N, T) = (200, 6)												
$\ln(w_3/w_1)$	0.6314	0.4391	−0.3838	0.3063	0.6342	0.4613	−0.3847	0.3769	0.6325	0.4458	−0.3842	0.3260
$[\ln(w_3/w_1)]^2$	−0.0003	0.0184	0.7112	0.6546	0.0004	0.0226	0.7145	0.7277	0.0001	0.0196	0.7125	0.6755
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	0.0003	0.0181	−0.5413	0.3237	−5.45E-06	0.0219	−0.5389	0.3356	0.0001	0.0192	−0.5404	0.3269
$\ln y_1\ \ln(w_3/w_1)$	−0.1877	0.0384	−0.9113	0.8611	−0.1886	0.0402	−0.9080	0.8685	−0.1880	0.0390	−0.9100	0.8626
$\ln(w_2/w_1)$	−0.6563	0.4722	0.1914	0.0661	−0.6566	0.4904	0.1884	0.0782	−0.6565	0.4774	0.1903	0.0695
$[\ln(w_2/w_1)]^2$	−0.0002	0.0201	0.7459	0.5689	−0.0005	0.0244	0.7469	0.5762	−0.0002	0.0213	0.7463	0.5712
$\ln y_1\ \ln(w_2/w_1)$	0.1948	0.0412	0.2014	0.0523	0.1951	0.0428	0.1998	0.0571	0.1949	0.0417	0.2008	0.0536
(N, T) = (200, 10)												
$\ln(w_3/w_1)$	0.5261	0.2911	−0.3955	0.2086	0.5274	0.3000	−0.3955	0.2355	0.5265	0.2936	−0.3955	0.2157
$[\ln(w_3/w_1)]^2$	0.0008	0.0083	0.7221	0.5714	0.0013	0.0099	0.7230	0.6002	0.0010	0.0088	0.7224	0.5792
$\ln(w_2/w_1)\ \ln(w_3/w_1)$	0.0003	0.0084	−0.5484	0.3112	−0.0005	0.0097	−0.5456	0.3135	0.0000	0.0087	−0.5473	0.3114
$\ln y_1\ \ln(w_3/w_1)$	−0.1563	0.0256	−0.9162	0.8497	−0.1566	0.0263	−0.9137	0.8503	−0.1564	0.0258	−0.9153	0.8494
$\ln(w_2/w_1)$	−0.5441	0.3102	0.1974	0.0488	−0.5444	0.3174	0.1946	0.0528	−0.5442	0.3122	0.1963	0.0498
$[\ln(w_2/w_1)]^2$	−0.0013	0.0094	0.7488	0.5648	−0.0006	0.0109	0.7493	0.5678	−0.0010	0.0098	0.7490	0.5657
$\ln y_1\ \ln(w_2/w_1)$	0.1615	0.0272	0.1989	0.0435	0.1617	0.0278	0.1982	0.0454	0.1616	0.0274	0.1986	0.0440
(N, T) = (200, 20)												
$\ln(w_3/w_1)$	0.4047	0.1674	−0.4005	0.1729	0.4054	0.1695	−0.4010	0.1792	0.4049	0.1680	−0.4007	0.1745

Table 3 continued

(σ^2, λ)	(1.88, 1.66)				(1.35, 0.83)				(1.63, 1.24)			
	Model A		Model B		Model A		Model B		Model A		Model B	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$[\ln(w_3/w_1)]^2$	−0.0007	0.0036	0.7215	0.5323	−0.0011	0.0039	0.7224	0.5395	−0.0008	0.0037	0.7218	0.5343
$\ln(w_2/w_1)$ $\ln(w_3/w_1)$	−4.81E-05	0.0037	−0.5495	0.3044	0.0004	0.0039	−0.5497	0.3056	0.0001	0.0037	−0.5496	0.3047
$\ln y_1 \ln(w_3/w_1)$	−0.1199	0.0147	−0.9208	0.8503	−0.1201	0.0148	−0.9209	0.8515	−0.1200	0.0147	−0.9209	0.8506
$\ln(w_2/w_1)$	−0.4240	0.1835	0.2002	0.0424	−0.4243	0.1852	0.2003	0.0434	−0.4241	0.1840	0.2002	0.0426
$[\ln(w_2/w_1)]^2$	0.0007	0.0039	0.7501	0.5636	0.0002	0.0042	0.7502	0.5642	0.0005	0.0040	0.7501	0.5638
$\ln y_1 \ln(w_2/w_1)$	0.1258	0.0161	0.1997	0.0408	0.1258	0.0162	0.1995	0.0412	0.4049	0.1680	−0.4007	0.1745

estimators of Models A and B. Does this imply that Step 3 is redundant? The answer is no. It is necessary for the estimation of the distribution parameters of v and U.

The distribution parameters are estimated in Step 5 by the maximum likelihood, and Table 4 presents the results. The estimators of Model C have larger biases and MSEs in comparison with those of Models A and B in most cases. We therefore drop Model C from now on whenever not necessary and focus our attention only on Models A and B. For the case of $(\sigma^2, \lambda) = (1.88, 1.66)$, despite the fact that Model B's estimator of γ has lower biases and MSEs than Model A does in almost all cases, the differences are quite small. Model B's estimator of σ^2 performs slightly better than Model A's, while the reverse is true for the estimator of λ. It is a caveat that Model A's estimator of σ^2 tends to have a larger variation when the sample size is small. As far as the estimator of smooth function $M(\cdot)$ is concerned, Model A is found to be superior to Model B since the former yields much smaller biases and MSEs than the latter does in most cases. Only for the cases of a large time period ($T = 20$) are Model B's biases a little less than Model A. Turning to the cases of $(\sigma^2, \lambda) = (1.35, 0.83)$ and $(1.63, 1.24)$, the results are rather similar to the preceding case.

Although both Models A and B perform reasonably well, the simulation results appear to be in favor of an advantage for Model A in general and for the estimation of TE scores in particular (see Table 6 below). Comparing (3-13) with (4-1), one can tell that their disparity originates from how β and $\ln G_{nt}$ are estimated. For Model A, they are estimated by NISUR viewing parameters contained in $\ln G_{nt}$ as unknown, while for Model B, $\ln G_{nt}$ is replaced by $\ln \hat{G}_{nt}$ leaving β to be estimated by OLS. The superiority of Model A may be explained by its allowance for the presence of $\ln G_{nt}$ in the expenditure equation.

It is apparent from Table 4 that Model C gives rise to undesirable estimators of $(\gamma, \sigma^2, \lambda)$. This is mainly ascribable to the fact that it overlooks Step 3 and proceeds from Step 2 directly to Steps 4 and 5. By doing so, the residual of (4-2) is indirectly derived using the NISUR estimates of $\ln \hat{G}_{nt}$ and $\hat{\beta}$, which are obtained by simultaneously estimating the $(J - 1)$ share equations, instead of the expenditure equation. Conversely, the residuals of (3-13) and (4-1) corresponding to Models A and B, respectively, are directly deduced from estimating the expenditure equation. Step 3 is thus necessary.

We have learned from Tables 2 and 3 that the parameter estimates of the parametric part of the cost function obtained in the first step outperform those obtained in the third step. These estimates are applied to compute $\ln G_{nt}$. We now compare the performance of the estimated $\ln G_{nt}$ to gain further insight into the properties of alternative models. Not surprisingly, G1 has smaller biases and MSEs than G2A and G2B, derived from Models A and B, respectively, in almost all (N, T) and (σ^2, λ) combinations. The outcomes support the use of G1 as the estimate of $\ln G_{nt}$. The simulation results are available upon request from the authors.

Table 5 shows the performance of the estimated AE measures. Again, the biases of AE1 are found to be smaller than both AE2A and AE2B, derived from Models A and B, respectively, in most (N, T) and (σ^2, λ) bundles, although the biases of both AE2A and AE2B measures are already tiny. In addition, those biases and MSEs are decreasing as the sample size grows. One is led to conclude that AE1 is superior to AE2A and AE2B as an estimate of the AE.

Given that technical efficiency often constitutes one of the primary issues in empirical studies, we thereby investigate the performance of TE measures based on the three models. It can be seen from Table 6 that the biases and MSEs of Model C vary dramatically, implying that the model is apt to yield uncertain and incredible TE measures. However, the biases and MSEs of both Models A and B are much less than those of Model C. The other two cases of (σ^2, λ) give similar implications. Comparing the first two columns of Table 6 with the middle two columns of the same table, we observe that the performance of the

Table 4 The performance of the estimators of (γ, σ^2, λ) setting $M(\cdot) = 2\ \ln(1 + y_1)$

	γ		σ^2		λ		$M(\cdot)$	
(*N*, *T*)	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(γ, σ^2, λ) = (0.025, 1.88, 1.66)								
Model A								
(100, 6)	0.0007	0.0006	−0.1253	0.1108	−0.0808	0.0432	0.0147	0.0071
(100, 10)	0.0007	0.0001	−0.1207	0.1032	−0.0975	0.0377	0.0098	0.0037
(100, 20)	0.0012	1.48E-05	−0.1498	0.1007	−0.1758	0.0507	0.0087	0.0024
(200, 6)	0.0009	0.0003	−0.0877	0.0557	−0.0568	0.0224	0.0093	0.0040
(200, 10)	0.0005	0.0001	−0.0845	0.0497	−0.0685	0.0181	0.0066	0.0023
(200, 20)	0.0011	8.00E − 06	−0.1233	0.0540	−0.1467	0.0312	0.0077	0.0019
Model B								
(100, 6)	0.0003	0.0007	−0.1101	0.1060	−0.0870	0.0433	0.1518	0.0363
(100, 10)	0.0004	0.0001	−0.1027	0.0979	−0.1046	0.0372	0.1073	0.0185
(100, 20)	0.0009	1.37E-05	−0.1307	0.0938	−0.1879	0.0530	0.0077	0.0039
(200, 6)	0.0006	0.0003	−0.0769	0.0535	−0.0598	0.0223	0.1472	0.0289
(200, 10)	0.0004	0.0001	−0.0729	0.0472	−0.0731	0.0181	0.1048	0.0152
(200, 20)	0.0010	7.45E-06	−0.1127	0.0506	−0.1552	0.0329	0.0067	0.0028
Model C								
(100, 6)	−0.4715	27.8117	0.5424	1.8874	−0.7704	0.9213	0.0787	0.2948
(100, 10)	−0.1906	11.8544	0.3473	0.8185	−0.8647	1.0488	0.0727	0.2177
(100, 20)	−0.0586	3.0532	0.1286	0.3725	−0.9295	1.1387	0.0438	0.1089
(200, 6)	−0.1163	6.2152	0.2694	0.4872	−0.6770	0.7413	0.0625	0.1903
(200, 10)	−0.0792	4.6876	0.1382	0.2180	−0.6909	0.7443	0.0329	0.1112
(200, 20)	−0.0365	1.3757	0.0279	0.1321	−0.8049	0.8860	0.0368	0.0573
(γ, σ^2, λ) = (0.025, 1.35, 0.83)								
Model A								
(100, 6)	0.0020	0.0037	−0.0705	0.0730	−0.0245	0.0278	0.0118	0.0100
(100, 10)	0.0009	0.0006	−0.0689	0.0643	−0.0281	0.0176	0.0070	0.0046
(100, 20)	0.0012	5.46E − 05	−0.0912	0.0560	−0.0511	0.0127	0.0061	0.0022
(200, 6)	0.0020	0.0018	−0.0501	0.0357	−0.0187	0.0145	0.0067	0.0054
(200, 10)	0.0005	0.0003	−0.0478	0.0306	−0.0194	0.0085	0.0043	0.0027
(200, 20)	0.0012	2.85E-05	−0.0737	0.0286	−0.0436	0.0068	0.0054	0.0016
Model B								
(100, 6)	0.0016	0.0037	−0.0647	0.0726	−0.0245	0.0279	0.1524	0.0403
(100, 10)	0.0005	0.0006	−0.0594	0.0629	−0.0270	0.0171	0.1108	0.0208
(100, 20)	0.0009	5.28E-05	−0.0792	0.0531	−0.0500	0.0119	0.0056	0.0039
(200, 6)	0.0016	0.0018	−0.0454	0.0355	−0.0180	0.0146	0.1497	0.0315
(200, 10)	0.0003	0.0003	−0.0413	0.0297	−0.0186	0.0083	0.10947	0.017091
(200, 20)	0.0010	2.77E-05	−0.0668	0.0272	−0.0430	0.0065	0.0046	0.0026
Model C								
(100, 6)	−0.8983	50.7877	0.6066	2.2731	−0.2619	0.2349	0.0408	0.2655
(100, 10)	−0.6136	42.2780	0.4063	0.9710	−0.3193	0.2484	0.0555	0.1863
(100, 20)	−0.1489	10.9556	0.1234	0.2132	−0.3376	0.1997	0.0260	0.0986
(200, 6)	−0.1513	7.1550	0.3134	0.5957	−0.2163	0.1378	0.0387	0.1582
(200, 10)	−0.0430	1.4204	0.1496	0.1566	−0.2203	0.1111	0.0160	0.0917
(200, 20)	0.0023	0.0051	0.0378	0.0537	−0.2761	0.1307	0.0216	0.0494
(γ, σ^2, λ) = (0.025, 1.63, 1.24)								
Model A								
(100, 6)	0.0010	0.0013	−0.1005	0.0906	−0.0471	0.0321	0.0132	0.0079

Table 4 continued

(N, T)	γ Bias	γ MSE	σ^2 Bias	σ^2 MSE	λ Bias	λ MSE	$M(\cdot)$ Bias	$M(\cdot)$ MSE
(100, 10)	0.0008	0.0002	−0.0972	0.0825	−0.0568	0.0245	0.0085	0.0039
(100, 20)	0.0012	2.48E-05	−0.1220	0.0774	−0.1020	0.0249	0.0075	0.0023
(200, 6)	0.0012	0.0006	−0.0708	0.0450	−0.0343	0.0167	0.0081	0.0689
(200, 10)	0.0006	0.0001	−0.0680	0.0394	−0.0400	0.0117	0.0056	0.0023
(200, 20)	0.0012	1.32E-05	−0.1000	0.0407	−0.0855	0.0143	0.0067	0.0017
Model B								
(100, 6)	0.0006	0.0013	−0.0892	0.0881	−0.0486	0.0319	0.1555	0.0386
(100, 10)	0.0004	0.0002	−0.0827	0.0791	−0.0579	0.0235	0.1108	0.2019
(100, 20)	0.0009	2.35E-05	−0.1058	0.0724	−0.1053	0.0242	0.0067	0.0038
(200, 6)	0.0008	0.0006	−0.0624	0.0438	−0.0345	0.0165	0.1518	0.0308
(200, 10)	0.0004	0.0001	−0.0586	0.0377	−0.0407	0.0114	0.10894	0.0163
(200, 20)	0.0010	1.26E-05	−0.0908	0.0383	−0.0880	0.0143	0.0058	0.0026
Model C								
(100, 6)	−1.1345	66.3433	0.3065	1.7909	−0.9243	1.0864	0.0648	0.2771
(100, 10)	−0.4090	27.9652	0.1410	0.9314	−0.9896	1.1737	0.0681	0.2019
(100, 20)	−0.1664	8.8257	−0.1376	0.2607	−1.0234	1.2005	0.0351	0.1006
(200, 6)	−0.1698	7.2282	0.0542	0.6162	−0.8542	0.9065	0.0498	0.1772
(200, 10)	−0.0697	2.3189	−0.1117	0.1607	−0.8494	0.8528	0.0201	0.0978
(200, 20)	0.1329	6.2066	−0.2256	0.1287	−0.9415	1.0193	0.0305	0.0568

Table 5 The performance of estimated AE setting $M(\cdot) = 2\ \ln(1 + y_1)$

			Model A		Model B	
	AE1		AE2A		AE2B	
(N, T)	Bias	MSE	Bias	MSE	Bias	MSE
$(\gamma, \sigma^2, \lambda) = (0.025, 1.88, 1.66)$						
(100, 6)	0.0004	0.0012	0.0023	0.0006	0.0017	0.0012
(100, 10)	0.0002	0.0007	0.0009	0.0004	0.0001	0.0007
(100, 20)	2.89E-06	0.0003	0.0008	0.0002	0.0001	0.0003
(200, 6)	0.0001	0.0006	0.0010	0.0003	0.0008	0.0006
(200, 10)	2.89E-06	0.0003	0.0006	0.0002	0.0001	0.0003
(200, 20)	0.0004	0.0002	0.0006	0.0001	0.0002	0.0002
$(\gamma, \sigma^2, \lambda) = (0.025, 1.35, 0.83)$						
(100, 6)	0.0004	0.0012	0.0024	0.0007	0.0024	0.0014
(100, 10)	0.0002	0.0007	0.0010	0.0004	0.0005	0.0007
(100, 20)	2.89E-06	0.0003	0.0008	0.0002	0.0002	0.0003
(200, 6)	0.0001	0.0006	0.0011	0.0003	0.0012	0.0007
(200, 10)	2.89E-06	0.0003	0.0007	0.0002	0.0004	0.0004
(200, 20)	0.0004	0.0002	0.0006	0.0001	0.0002	0.0002
$(\gamma, \sigma^2, \lambda) = (0.025, 1.63, 1.24)$						
(100, 6)	0.0004	0.0012	0.0023	0.0006	0.0019	0.0013
(100, 10)	0.0002	0.0007	0.0010	0.0004	0.0002	0.0007
(100, 20)	2.89E-06	0.0003	0.0008	0.0002	0.0001	0.0003
(200, 6)	0.0001	0.0006	0.0011	0.0003	0.0009	0.0006
(200, 10)	2.89E-06	0.0003	0.0006	0.0002	0.0002	0.0003
(200, 20)	0.0004	0.0002	0.0006	0.0001	0.0002	0.0002

Table 6 The performance of estimated TE scores setting $M(\cdot) = 2\ \ln(1 + y_1)$

(N, T)	Model A		Model B		Model C	
	Bias	MSE	Bias	MSE	Bias	MSE
$(\gamma, \sigma^2, \lambda) = (0.025, 1.88, 1.66)$						
(100, 6)	0.0022	0.0062	0.0704	0.0118	39.7962	5.79E + 08
(100, 10)	0.0029	0.0027	0.0607	0.0069	0.0881	825.1991
(100, 20)	0.0204	0.0016	0.0204	0.0016	0.0373	0.0197
(200, 6)	0.0001	0.0059	0.0679	0.0110	0.0824	693.1477
(200, 10)	0.0012	0.0025	0.0589	0.0064	1.0285	1.94E + 06
(200, 20)	0.0200	0.0016	0.0200	0.0015	0.0314	0.0119
$(\gamma, \sigma^2, \lambda) = (0.025, 1.35, 0.83)$						
(100, 6)	−0.0007	0.0121	0.0522	0.0149	9149.6239	3.47E + 13
(100, 10)	0.0005	0.0059	0.0510	0.0087	0.0781	674.3316
(100, 20)	0.0175	0.0020	0.0177	0.0020	0.0274	0.0134
(200, 6)	−0.0034	0.0114	0.0498	0.0140	18.5228	3.88E + 08
(200, 10)	−0.0012	0.0056	0.0496	0.0083	0.0115	0.0175
(200, 20)	0.0172	0.0019	0.0173	0.0019	0.0223	0.0105
$(\gamma, \sigma^2, \lambda) = (0.025, 1.63, 1.24)$						
(100, 6)	0.0006	0.0085	0.0654	0.0133	6106.3710	2.03E + 13
(100, 10)	0.0019	0.0038	0.0589	0.0077	0.0644	301.2878
(100, 20)	0.0192	0.0017	0.0192	0.0017	0.0303	0.0135
(200, 6)	−0.0015	0.0081	0.0632	0.0125	53.6804	3.33E + 09
(200, 10)	0.0002	0.0036	0.0573	0.0073	0.0245	108.7752
(200, 20)	0.0188	0.0017	0.0188	0.0016	0.0270	0.0088

estimated TE score from Model A surpasses that of Model B in most (N, T) and (σ^2, λ) bundles. Model B may be valid only when T is greater than or equal to 20.[9]

We also conduct simulations assuming $\gamma = -0.025$ and the remaining parameter values are held intact. A negative value of γ implies that the TE score of a firm deteriorates over time. The results are similar to the foregoing and are not shown to save space. To understand the effects of dimensionality of smooth function $M(\cdot)$ on the performance of the various estimators, we re-specify M as a function of two variables, i.e., $\ln y_1$ and $\ln y_2$. Evidence is found that the performance of the estimators under consideration seems to be irrespective of the increase in the dimension of M and the explanatory variables in the parametric part. We further check the performance of our proposed estimator in the context of cross sectional data. Evidence is found that Model A performs at least as well as Model B, while Model C acts badly. The results do not change the previous findings except that TE scores of firms cannot be consistently estimated, since the variance of the conditional mean or the conditional mode for each individual firm does not vanish as the size of the cross section increases.

[9] We agree with the referee's opinion that measures of scale economies (SE) and cost elasticity (CE) are important topics particularly in conventional performance analysis. Evidence is found that the simulated estimates of the SE would accurately predict the scope of the true SE and the predictability rises as the sample size increases. Since the CE of outputs is the reciprocal of SE, its measure has very similar performance to the SE. Viewed from this angle, our modeling appears to provide satisfactory results.

6 Conclusions

Since most economic relationships predicted by economic theory are unknown, one has to count on a particular parametric form, which may lead to a biased estimation due to invalid model specification. The importance of nonparametric and semiparametric regression techniques has drawn much attention from econometricians and applied researchers recently. These techniques allow the functional form to be determined at least partially by the data. Fan et al. (1996) and Deng and Huang (2008) generalized the conventional linear stochastic frontier model to a semiparametric stochastic production frontier model. On the basis of previous works, this article adds to the current literature by considering both TE and AE in the context of a semiparametric stochastic cost frontier model using panel data.

This paper intends to solve two major problems faced by applied researchers. First, the cost system must be estimated simultaneously, suffering from computational difficulties. Second, the log-likelihood function of the expenditure equation cannot be maximized due to the presence of the nonparametric component. Even worse, the nonparametric function is unable to be estimated by existing nonparametric regression methods. We propose a five-step procedure to cope with these problems. Evidence from a set of Monte Carlo simulations tends to support the superiority of Model A at least for a moderate sample size, while the performance of Model B's estimators is nearly as good as that of Model A's, particularly when the time period of the panel data is long. In other words, Model B is appropriate for long panel data.

The first step estimators of the cost share equations perform reasonably well. We thus advocate using these estimates to compute the AE measure and treat the estimated allocative parameters as given in the following steps. It is noticeable that despite the uselessness of the parameter estimates obtained in the third step, this step is necessary to yield the residuals and to concentrate out variance σ^2. Otherwise, estimators of Step 5 will perform poorly. Moreover, Models A and B are robust to the inclusion of additional explanatory variables for both parametric and nonparametric portions of the cost function. When cross sectional data are available, the foregoing conclusions continue to hold in general, except that the bias of the estimated TE measure does not decrease as the sample size increases.

References

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econom 6:21–37

Akhigbe A, McNulty JE (2003) The profit efficiency of small US commercial banks. J Bank Finance 27:307–325

Altunbas Y, Evans L, Molyneux P (2001) Bank ownership and efficiency. J Money Credit Bank 33:926–954

Atkinson SE, Cornwell C (1993) Estimation of technical efficiency with panel data: A dual approach. J Econom 59:257–262

Atkinson SE, Cornwell C (1994) Parametric estimation of technical and allocative inefficiency with panel data. Int Econ Rev 35: 231–243

Battese GE, Coelli TJ (1992) Frontier production functions, technical efficiency and panel data:with application to paddy farmers in India. J Prod Anal 3:153–169

Bauer PW (1990) Recent development in the econometric estimation of frontiers. J Econ 46:39–56

Berger AN, DeYoung R (1997) Problem loans and cost efficiency in commercial banks. J Bank Finance 21:849–870

Berger AN, Mester LJ (1997) Inside the black box: what explains differences in the efficiencies of financial institutions? J Bank Finance 21:895–947

Berger AN, Hancock GA, Humphrey DB (1993) Bank efficiency derived from the profit function. J Bank Finance 17:317–347

Berger AN, Leusner JH, Mingo JJ (1997) The efficiency of bank branches. J Monet Econ 40:141–162

Deng WS, Huang TH (2008) A semiparametric approach to the estimation of the stochastic frontier model with time-variant technical efficiency. Acad Econ Pap 36:167–193

Fan Y, Li Q (1992) The asymptotic expansion of kernel sum of square residuals and its application in hypotheses testing, Discussion paper, University of Windsor, Department of Economics

Fan Y, Li Q, Weersink A (1996) Semiparametric estimation of stochastic production frontier models. J Bus Econ Stat 14: 460–468

Ferrier GD, Lovell CAK (1990) Measuring cost efficiency in banking. J Econom 46:229–245

Härdle W (1990) Applied nonparametric regression. Cambridge University Press, Cambridge

Härdle W, Stoker TM (1989) Investigating smooth multiple regression by the method of average derivatives. J Am Stat Assoc 84:986–995

Huang TH (2000) Estimating X-efficiency in Taiwanese banking using a Translog shadow profit function. J Prod Anal 14:225–245

Huang TH, Wang MH (2004) Comparisons of economic inefficiency between output and input measures of technical inefficiency using Fourier flexible cost frontiers. J Prod Anal 22:123–142

Huang TH, Shen CH, Chen KC, Tseng SJ (2011) Measuring technical and allocative efficiencies for banks in the transition countries using the Fourier flexible cost function. J Prod Anal 35:143–157

Kumbhakar SC (1991) The measurement and decomposition of cost inefficiency: The translog cost system. Oxf Econ Pap 43:667–683

Kumbhakar SC (1996a) A parametric approach to efficiency measurement using a flexible profit function. Southern Econ J 63:473–487

Kumbhakar SC (1996b) Efficiency measurement with multiple outputs and multiple inputs. J Prod Anal 7:225–256

Kumbhakar SC (1997) Modeling allocative inefficiency in a translog cost function and cost share equations: an exact relationship. J Econom 76:351–356

Kumbhakar SC, Lovell CAK (2000) Stochastic frontier analysis. Cambridge University Press, Cambridge

Kumbhakar SC, Wang HJ (2006a) Pitfalls in the estimation of a cost function that ignores allocative inefficiency: A Monte Carlo analysis. J Econom 134:317–340

Kumbhakar SC, Wang HJ (2006b) Estimation of technical and allocative inefficiency: a primal system approach. J Econom 134:419–440

Meeusen W, Van Den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18:435–444

Nadaraya EA (1964) On estimating regression. Theory Prob Appl 10:186–190

Olson JA, Schmidt P, Waldman DM (1980) A Monte Carlo study of estimators of stochastic frontier production. J Econom 13:67–82

Robinson PM (1988) Root-N-Consistent semiparametric regression. Econometrica 56:931–954

Wand MP, Jones MC (1995) Kernel smoothing. Chapman and Hall Press, New York

Watson GS (1964) Smooth regression analysis. Sankhya Series A 26:359–372

Wheelock DC, Wilson PW (2001) New evidence on returns to scale and product mix among US commercial banks. J Monet Econ 47:653–674

Yatchew A (1998) Nonparametric regression techniques in economics. J Econ Literature 36:669–721

Yatchew A (2003) Semiparametric regression for the applied econometrician. Cambridge University Press, Cambridge